The spectral computation of birth and death chains

Guan-Yu Chen

Dept. of Applied Math, NCTU

April 9, 2013 at Providence University

1. Mixing time for Markov chains
2. Separation mixing of birth and death chains
3. Computing the spectral gap
4. Computing the whole spectrum

Fundamental setting

- $\mathcal{X}$: a finite set.
- K: a matrix indexed by $\mathcal{X}$ satisfying

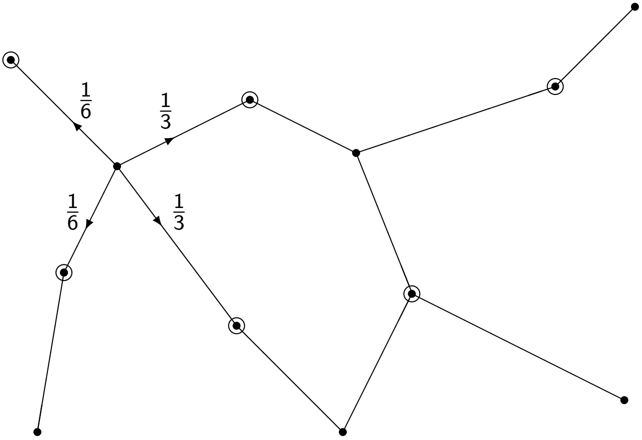
$$K(x,y) \geq 0, \quad \sum_y K(x,y) = 1.$$

- X_m: a (discrete time) Markov chain on $\mathcal{X}$ with transition matrix K. That is,

$$\mathbb{P}(X_{m+1} = y | X_m = x, X_{m-1} = i_{m-1}, ..., X_0 = i_0) = K(x,y).$$

- K is irreducible if, for any $x, y \in \mathcal{X}$, there is $m > 0$ such that $K^m(x,y) > 0$.
- If K is irreducible, then there is a unique probability π on $\mathcal{X}$ satisfying $\pi K = \pi$. π is called the stationary distribution of K.

Random walks on graphs



Converging to the stationarity

Theorem

Let K be irreducible with stationary distribution π. If K is aperiodic, i.e. there is $m > 0$ such that $K^m(x,y) > 0$ for all $x, y \in \mathcal{X}$, then

$$\lim_{m\to\infty} \mathbb{P}(X_m = y | X_0 = x) = \lim_{m\to\infty} K^m(x,y) = \pi(y), \quad \forall x, y \in \mathcal{X}.$$

Measuring the convergence

Let μ be the distribution of X_0.

- Total variation:

$$d_{\text{TV}}(\mu, m) = \|\mu K^m - \pi\|_{\text{TV}} := \max_{A \subset \mathcal{X}} |\mu K^m(A) - \pi(A)|.$$

- Separation:

$$d_{\text{sep}}(\mu, m) = \text{sep}_\pi(\mu K^m) := \max_{y \in \mathcal{X}} \{1 - \mu K^m(y)/\pi(y)\}.$$

- Mixing time: Let d be any above distance. The mixing time in d is defined by

$$T_d(\mu, \epsilon) = \min\{m \geq 0 | d(\mu, m) \leq \epsilon\}, \quad \forall \epsilon > 0.$$

Birth and death chains

Let $n \in \mathbb{N}$.

- $\mathcal{X} = \{0, 1, ..., n\}$.
- Birth rate p_i, death rate q_i and holding rate r_i satisfying

$$p_i + q_i + r_i = 1, \quad p_n = q_0 = 0.$$

- A birth and death chain is a Markov chain on $\mathcal{X}$ with transition matrix K given by

$$K(i, i+1) = p_i, \quad K(i, i-1) = q_i, \quad K(i, i) = r_i.$$

- K is irreducible if $p_i q_{i+1} > 0$ for $0 \le i < n$.
- If K is irreducible, then K is aperiodic if and only if $r_i > 0$ for some i.
- If K is irreducible with stationary distribution π, then

$$\pi(i) = c \frac{p_0 p_1 \cdots p_{i-1}}{q_1 q_2 \cdots q_i},$$

where c is a normalizing constant.

Separation distance

Theorem (Diaconis & Saloff-Coste(2006))

Let K be the transition matrix of an irreducible birth and death chain on $\{0,1,...,n\}$. Let $\lambda_1,...,\lambda_n$ be nonzero eigenvalues of $I-K$. Suppose $p_i+q_{i+1}\le 1$ for $0\le i<n$. Then,

$$d_{\text{sep}}(\delta_0,m)=d_{\text{sep}}(\delta_n,m)=\sum_{i=1}^{n}\prod_{j\neq i}\frac{\lambda_j}{\lambda_j-\lambda_i}(1-\lambda_i)^m,$$

where δ_i is the Dirac measure concentrated on i.

Separation distance

Refer to the setting in the above Theorem.

(1) Let T be a random variable satisfying

$$\mathbb{P}(T > m) = d_{\text{sep}}(\delta_0, m).$$

Then,

$$\mathbb{E}T = \sum_{i=1}^{n} \frac{1}{\lambda_i}, \quad \text{Var}(T) = \sum_{i=1}^{n} \frac{1-\lambda_i}{\lambda_i^2}.$$

(2) If $\lambda_i < 1$ for $1 \le i \le n$, then T is the sum of independent geometric random variables with success probabilities $\lambda_1, \ldots, \lambda_n$.

Separation mixing time

Recall the one-sided Chebyshev inequality. Let $\mathrm{SD}(Y)$ be the standard deviation of a random variable Y. For $a > 0$,

$$\mathbb{P}(Y > \mathbb{E}Y + a\mathrm{SD}(Y)) \le \frac{1}{1+a^2}, \quad \mathbb{P}(Y < \mathbb{E}Y - a\mathrm{SD}(Y)) \le \frac{1}{1+a^2}.$$

This implies, for $\epsilon \in (0,1)$,

$$\lfloor \mathbb{E}T - (1/\epsilon - 1)^{-1/2}\mathrm{SD}(T) \rfloor \le T_{\text{sep}}(\delta_0, \epsilon) \le \lceil \mathbb{E}T + (1/\epsilon - 1)^{1/2}\mathrm{SD}(T) \rceil,$$

where $\lfloor t \rfloor = \max\{z \in \mathbb{Z} | z \le t\}$ and $\lceil t \rceil = \min\{z \in \mathbb{Z} | z \ge t\}$.

Ehrenfest chains

- A lazy Ehrenfest chain on $\{0, 1, ..., n\}$ is a birth and death chain with

$$p_i = 1/2 - i/(2n), \quad q_i = i/(2n), \quad r_i = 1/2.$$

- The stationary distribution π satisfies $\pi(i) = \binom{n}{i} 2^{-n}$.
- The transition matrix of the Ehrenfest chain has eigenvalues $\beta_i = 1 - i/n$ for $0 \le i \le n$ and eigenvectors

$$\psi_i(j) = \binom{n}{i}^{-1/2} \sum_{k=0}^{i} (-1)^k \binom{j}{k} \binom{n-j}{i-k},$$

where ψ_i is related to Krawtchouk polynomials. (Difference relation)

- Clearly, $\mathbb{E}T = n \log n + O(1)$, $\mathrm{SD}(T) \asymp n$ and

$$T_{\text{sep}}(\delta_0, \epsilon) = n \log n + O(n), \quad \forall \epsilon \in (0, 1).$$

Qualitative behavior

Theorem

Let K be the transition matrix of an irreducible birth and death chain on $\{0, 1, ..., n\}$ with rates p_i, q_i, r_i. Let β be an eigenvalue of K with associated eigenvector ψ. Then, for $0 \leq i \leq n$,

$$\beta\psi(i) = p_i\psi(i+1) + q_i\psi(i-1) + r_i\psi(i), \tag{1}$$

where $p_n = q_0 = 0$. Furthermore, if β is the second largest eigenvalue of K and $\psi(0) < 0$, then

$$\psi(i) < \psi(i+1), \quad \forall 0 \leq i < n.$$

Remark

The monotonicity of ψ due to the work of Miclo in 2009.

Qualitative behavior

- If ψ is an eigenvector associated with a non-zero eigenvalue of $I-K$, then $\pi(\psi)=0$.
- Subtracting $\psi(i)$ from both sides of (1) gives

$$\psi(i+1)=\psi(i)+\frac{[\psi(i)-\psi(i-1)]q_i-\lambda\psi(i)}{p_i},$$

  where $\lambda=1-\beta$.
- If β is the second largest eigenvalue of K, then λ is called the spectral gap of K. Furthermore,

$$\lambda=\inf\left\{\frac{\mathcal{E}(f,f)}{\mathrm{Var}_\pi(f)}\middle| f\text{ is non-constant}\right\},$$

  where $\mathcal{E}(f,g):=\langle f,(I-K)g\rangle_\pi$ and $\mathrm{Var}_\pi(f)=\pi(f^2)-(\pi(f))^2$.

Characterizing the spectral gap

Theorem (Chen & Saloff-Coste(2012))

Let λ be the spectral gap of K. For $a > 0$, set $\phi_a(0) = -1$ and

$$\phi_a(i+1) = \phi_a(i) + \frac{\{[\phi_a(i) - \phi_a(i-1)]q_i - a\phi_a(i)\}^+}{p_i},$$

where $t^+ = \max\{t, 0\}$. Set $L(a) = \mathcal{E}(\phi_a, \phi_a)/\mathrm{Var}_\pi(\phi_a)$. Then,

$$L(a) = a \quad \Leftrightarrow \quad \pi(\phi_a) = 0 \quad \Leftrightarrow \quad a = \lambda.$$

Furthermore,

$$\pi(\phi_a) \begin{cases} > 0 & \text{for } a > \lambda, \\ < 0 & \text{for } 0 < a < \lambda. \end{cases}$$

Algorithm 1

Theorem (Chen & Saloff-Coste (2012))

Let L, λ be as in the previous theorem. Fix $a > 0$ and set $\lambda_0 = a$ and $\lambda_{k+1} = L(\lambda_k)$. For $n \geq 2$:

(1) If $a = \lambda$, then $\lambda_k = \lambda$ for all k.

(2) If $a \neq \lambda$, then $\lambda_k > \lambda_{k+1} > \lambda$.

(3) There is $\sigma \in (0, 1)$ such that

$$0 \leq \lambda_k - \lambda \leq \sigma^{k-1}\lambda_1, \quad \forall k \geq 1.$$

Remark

For $n = 1$, $\mathcal{E}(f, f) = \lambda \mathrm{Var}_\pi(f)$ for any f.

Algorithm 2

- Set $L_0 = 0$ and $U_0 = 2$.
- For $k = 0, 1, ...,$ define $\lambda_k = (L_k + U_k)/2$ and

$$\begin{cases} L_{k+1} = L_k, U_{k+1} = \lambda_k & \text{if } \pi(\phi_{\lambda_k}) > 0 \\ L_{k+1} = \lambda_k, U_{k+1} = U_k & \text{if } \pi(\phi_{\lambda_k}) < 0, \\ L_{k+1} = U_{k+1} = \lambda_k & \text{if } \pi(\phi_{\lambda_k}) = 0. \end{cases}$$

Theorem

Referring to the above setting, one has

$$0 \leq \max\{U_k - \lambda, \lambda - L_k\} \leq 2^{1-k}, \quad \forall k \geq 0.$$

Qualitative behavior

Theorem (L. Miclo (2008))

Let K be the transition matrix of a birth and death chain on $\{0, 1, ..., n\}$. For $1 \le i \le n$, let λ_i be the i-th smallest non-zero eigenvalue of $I - K$ and ζ_i be an eigenvalue associated with λ_i satisfying $\zeta_i(1) < 0$. Then, there are

$$0 = a_1 < b_1 \le a_2 < b_2 \le \cdots \le a_i < b_i = n$$

satisfying $a_{j+1} - b_j \in \{0, 1\}$ such that $\zeta_i(b_j) = \zeta_i(a_{j+1})$ and

$$\zeta_i \text{ is strictly } \begin{cases} \text{increasing on } [a_j, b_j] & \text{if } j \text{ is odd,} \\ \text{decreasing on } [a_j, b_j] & \text{if } j \text{ is even.} \end{cases}$$

Vectors of Type i

Definition

A function $f : \{0, 1, ..., n\} \to \mathbb{R}$ is called "Type i" if there are

$$0 = a_1 < b_1 \leq a_2 < b_2 \leq \cdots \leq a_i < b_i \leq n$$

satisfying $a_{j+1} - b_j \in \{0, 1\}$ such that $f(b_j) = f(a_{j+1})$ and

(1) f is strictly monotonic on $[a_j, b_j]$ for all $1 \leq j \leq i$.

(2) $[f(a_j) - f(a_j + 1)][f(a_{j+1}) - f(a_{j+1} + 1)] < 0$.

(3) $f(k) = f(b_i)$ for $b_i \leq k \leq n$.

The points a_j, b_j are called "peak-valley points" of f.

Vectors of Type i

- For $a > 0$, let $\phi_a(0) = -1$ and set

$$\phi_a(i+1) = \phi_a(i) + \frac{[\phi_a(i) - \phi_a(i-1)]q_i - a\phi_a(i)}{p_i}.$$

- Suppose ϕ_a is of type i with peak-valley points $a_1, b_1, ..., a_i, b_i$. For $1 \le j < i$, define

$$\phi_a^{(j)}(k) = \begin{cases} \phi_a(k) & \forall 0 \le k \le b_j, \\ \phi_a(b_j) & \forall b_j < k \le n. \end{cases}$$

  For $i \le j \le n$, set $\phi_a^{(j)} = \phi_a$.

- For $a > 0$, set

$$\mathcal{L}(a) = \frac{\mathcal{E}(\phi_a, \phi_a)}{\text{Var}_\pi(\phi_a)}, \quad \mathcal{L}^{(j)}(a) = \frac{\mathcal{E}(\phi_a^{(j)}, \phi_a^{(j)})}{\text{Var}_\pi(\phi_a^{(j)})}$$

Local convergence

Theorem (Chen & Saloff-Coste (2012))

Let $\lambda_1 < \cdots < \lambda_n$ be the non-zero eigenvalues of $I - K$. Then, there is $\epsilon > 0$ such that, for $1 \le i \le n$ and $|a - \lambda_i| < \epsilon$,

(1) $\phi_a = \phi_a^{(i)}$,

(2) Set $a_0 = a$ and $a_{k+1} = \mathcal{L}(a_k)$. Then, $(a_k)_{k=1}^{\infty}$ is monotonic and converges to λ_i.

Characterizing the spectrum

Theorem (Chen & Saloff-Coste (2012))

Let $\lambda_1 < \cdots < \lambda_n$ be non-zero eigenvalues of $I - K$ and ϕ_a be the vector defined as before. Then, there are constants $\alpha_i \in (\lambda_i, \lambda_{i+1})$ with $1 \le i \le n-1$ such that

(1) ϕ_a is of type i for $a \in (\alpha_{i-1}, \alpha_i]$, where $a_0 := 0$ and $a_n := \infty$.

(2) $\pi(\phi_a) > 0$ for $a \in (\lambda_{2i-1}, \lambda_{2i})$, where $\lambda_0 := 0$ and $\lambda_{n+1} := \infty$.

(3) $\pi(\phi_a) < 0$ for $a \in (\lambda_{2i}, \lambda_{2i+1})$.

Algorithm computing the ith smallest eigenvalue

- Set $L_0 = 0$ and $U_0 = 2$.
- For $k = 0, 1, \ldots$, let $\lambda_k = (L_k + U_k)/2$ and j be the type of ϕ_{λ_k}. Set

$$\begin{cases} L_{k+1} = L_k, U_{k+1} = \lambda_k & \text{if } j > i \text{ or if } j = i,\ (-1)^{i+1}\pi(\phi_{\lambda_k}) > 0, \\ L_{k+1} = \lambda_k, U_{k+1} = U_k & \text{if } j < i \text{ or if } j = i,\ (-1)^{i+1}\pi(\phi_{\lambda_k}) < 0, \\ L_{k+1} = U_{k+1} = \lambda_k & \text{if } j = i,\ \pi(\phi_{\lambda_k}) = 0. \end{cases}$$

Theorem

Let λ_i be the ith smallest non-zero eigenvalue of $I - K$. Then,

$$0 \le \max\{U_k - \lambda_i, \lambda_i - L_k\} \le 2^{1-k}, \quad \forall k \ge 0.$$

Reference

1. Chen, G.-Y. and Saloff-Coste, L. *The L^2-cutoff for reversible Markov processes.* Journal of Functional Analysis, 258, 2246–2315, 2010.
2. Chen, G.-Y. and Saloff-Coste, L. *Spectral computations of birth and death chains.* In revision. 2012
3. Diaconis, P. *Group representations in probability and statistics,* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 1988.
4. Diaconis, P. and Saloff-Coste, L. *Separation cutoffs for birth and death chains.* The Annals of Applied Probability, 16, 2098–2122, 2006.
5. Saloff-Coste, L. *Lectures on finite Markov chains*, Lectures on probability theory and statistics (Saint-Flour), 1996.

Thank you!